DOCUMENT RESUME

ED 303 490 TM 012 705

AUTHOR DuRapau, Theresa M.
TITLE Benefits of Using Planned Comparisons Rather Than Post Hoc Tests: A Brief Review with Examples.
PUB DATE Nov 88
NOTE 15p.; Paper presented at the Annual Meeting of the Mid-South Educational Research Association (17th, Louisville, KY, November 9-11, 1988).
PUB TYPE Information Analyses (070) -- Speeches/Conference Papers (150)

EDRS PRICE MF01/PC01 Plus Postage.
DESCRIPTORS Analysis of Covariance; *Analysis of Variance; *Comparative Analysis; *Error of Measurement; Literature Reviews; *Multivariate Analysis
IDENTIFIERS *A Priori Tests; *Planned Comparisons; Post Hoc Tests; Type II Errors

ABSTRACT

The rationale behind analysis of variance (including analysis of covariance and multiple analyses of variance and covariance) methods is reviewed, and unplanned and planned methods of evaluating differences between means are briefly described. Two advantages of using planned or a priori tests over unplanned or post hoc tests are presented. In addition to increasing power against Type II error rates, planned comparisons force the researcher to plan the research design more thoughtfully. An example illustrating the greater statistical power of planned comparisons is also provided. Three data tables are included, and two figures are appended. (Author/TJH)

****
* Reproductions supplied by EDRS are the best that can be made *
* from the original document. *
****

ED303490

BENEFITS OF USING PLANNED COMPARISONS RATHER THAN POST HOC TESTS: A BRIEF REVIEW WITH EXAMPLES

U.S. DEPARTMENT OF EDUCATION
Office of Educational Research and Improvement
EDUCATIONAL RESOURCES INFORMATION CENTER (ERIC)

☑ This document has been reproduced as received from the person or organization originating it.

☐ Minor changes have been made to improve reproduction quality.

• Points of view or opinions stated in this document do not necessarily represent official OERI position or policy.

Theresa M. DuRapau, R.S.M.

University of New Orleans

"PERMISSION TO REPRODUCE THIS MATERIAL HAS BEEN GRANTED BY
THERESA M. DURAPAU

TO THE EDUCATIONAL RESOURCES INFORMATION CENTER (ERIC)."

Paper presented at the annual meeting of the Mid-South Educational Research Association, Louisville, KY, November 9, 1988.

TM012 705

ABSTRACT

The rationale behind OVA methods is reviewed, followed by a brief explanation of unplanned and planned methods of evaluating differences between means. Two advantages of using planned or a priori tests over unplanned or post hoc tests are presented. In addition to increasing power against Type II error rates, planned comparisons force the researcher to more thoughtfully plan the research design. An example illustrating the greater statistical power of planned comparisons is presented.

The evaluation of differences between means is a common statistical exercise. The statistical significance of differences between groups of means is most frequently determined through the analysis of variance (ANOVA) methods introduced by Fisher (1925). Other ANOVA-type methods (ANCOVA, MANOVA, MANCOVA) are also used and are henceforth referred to here as OVA methods.

OVA methods test the null hypothesis of equality among means. Where there are more than k = 2 groups and the null hypothesis is rejected, the question then becomes: where are the differences? Moore (1983, p. 299) makes the suggestion that:

> If we have statistical significance when we have only two groups, and thus only two means, we can visually inspect the data to determine which group performed better than the other. But when we have three or more groups, we need to investigate specific mean comparisons.

Ways of Isolating Statistically Significantly Different Means

In an effort to isolate means that are statistically significantly different, researchers may make use of one of two types of multiple comparisons: unplanned comparisons (referred to as **a posteriori** or **post hoc** tests and performed after a significant F is obtained for an omnibus null), and planned comparisons (referred to as **a priori** or **focused** tests and used in the place of an overall ANOVA). Due to the inflation of Type I error rate resulting from multiple

t-tests performed on all possible pairs of means, this particular procedure is unacceptable. The rationale for such a conclusion rests in the concept of experimentwise Type I error rate, which discussion is beyond the scope of this paper but which is explained elsewhere (Thompson, 1987).

Unplanned comparisons (post hoc tests). One method for isolating means that are statistically significantly different involves unplanned tests. These post hoc tests include the Scheffe, Tukey, and Duncan multiple comparison tests. Post hoc tests are employed after the achieving of a significant F test. Interest in determining where differences between means occur is developed "after the fact" (post hoc) of a significant omnibus, over-all, F test. Kirk (1984, p. 360) refers to this process as "ferreting out significant differences among means, or, as it is often called, data snooping."

Post hoc tests of comparisons among pairs of means usually involve performing all possible comparisons. Acknowledging the fact that some thought given to the experiment prior to its implementation by the researcher would eliminate unnecessary comparisons, post hoc tests are seen by some researchers as insufficient for the purpose for which they are performed. These tests also result in the inflation of the experimentwise Type I error rate.

In order to avoid the heightened posssibility of Type I errors, statistical adjustments (Games, 1971a, 1971b) are incorporated in unplanned tests of comparisons. Horvath (1985, p. 223) maintains that

since the performing of multiple comparisons within the context of a single study increases the probability of a Type I error, referred to as inflating experimentwise error rate, then "post F-test procedures must include some accommodation for this danger." However, lessening of power against Type II errors is the trade off resulting from adjusting for the inflation of the Type I error rate. Due to this trade-off, researchers look at a second method of isolating statistically different means.

Planned comparisons (a priori tests). Planned comparisons offer an alternative to unplanned comparisons and involve the weighting of data by **sets of contrasts.** The contrasts may be non-trend or trend (e.g. linear, quadratic, cubic) (Fisher & Yates, 1957) depending upon the nature of the research design. A data set for the example of non-trend contrasts in a six-level, one-way case was generated by Thompson (1987) and involves the dependent variable data, presented in Table 1 in the column headed DV.

INSERT TABLE 1 ABOUT HERE

The number of coding columns is equal to the number of degrees of freedom, df = k - 1, where k represents the number of levels in the way. That is, for a main effect there is one less coding contrast than the number of levels in the way. Each coding column depicts a contrast which represents a null hypothesis. Hays (1981, p. 425) notes

> each and every degree of freedom associated with treatments in any fixed-effects analysis of variance corresponds to some possible comparison of means. The number of degrees of freedom for the mean square between is the number of possible <u>independent</u> [i.e., orthogonal] comparisons to be made on the means.

The same information is represented in the coding columns as is involved in testing omnibus effects in the OVA analysis. The only difference is that coding results in the partitioning of the total sum of squares, SOS, of the dependent variable into smaller parts using separate one df hypotheses (Thompson, 1985).

Contrasts may be orthogonal or not. Winer (1971, p. 175) maintained that "whether these comparisons are orthogonal or not makes little or no difference." However, orthogonal planned comparisons do have a special appeal. Keppel (1982, p. 147) explains that orthogonal comparisons are uncorrelated. That is, decisions regarding the null hypothesis of one comparison (contrast, coding column) is not influenced by decisions regarding any other orthogonal comparison. Such independence seems desirable in the analysis of relationships among means. Contrasts are uncorrelated (orthogonal) when each contrast (coding column) totals zero and when the sum of the cross-products of each pair of contrasts total zero.

Reasons for Preferring Planned Comparisons

Two reasons are proposed for researchers' preferring planned comparisons over unplanned comparisons (Thompson, 1988). First, planned comparisons guarantee more power against Type II errors; that is, real differences among means are more likely to be detected. Pedhauzur (1982) maintains that planned comparisons among means are more powerful than post hoc comparisons; that is, "it is possible for a specific comparison to be not significant when tested by post hoc methods but significant when tested by a priori methods" (Pedhazur, 1082, pp. 304-305). The data analysis example presented in this paper illustrates this possibility. Glasnapp and Poggio (1985) encourage the use of planned comparisons when designing research using ANOVA because of the additional statistical power in a priori comparisons testing.

The second, and more important, reason for the benefit of planned comparisons is that they encourage the researcher to carefully think through the research project during the design stage. Keppel (1982) maintains that planned comparisons become the "motivating force" behind the experiment, requiring careful thought about the manipulation in an experimental study. Carefully thinking through the design of the research effort is at the core of planned comparisons. This is in contradistinction to the relatively haphazard development of the research design which does not focus on particular comparisons between means until after achieving a significant omnibus in the

overall F test. The latter inquiry is done post hoc after the discovery of a significant omnibus F.

Example

The data in Table 1 can be analyzed to illustrate these points in a concrete fashion. Notice should be taken of the fact that there are five coding columns (one less than the number of levels in the way) each totaling zero and the sums of their cross-products also equaling zero. Thus these contrasts are orthogonal. Since the design is a one-way case, there is no interaction and thus no interaction coding columns.

INSERT TABLE 2 ABOUT HERE

The researcher may be interested in using the analysis of variance in order to determine if any differences exist among means of groups. Table 2 presents the results of a one-way ANOVA of these data. The omnibus test was not statistically significant. Therefore the null hypothesis of equality among all six means is not rejected, indicating no significant statistical difference using the overall omnibus F test.

Contrary to traditional practice, a post hoc test could be conducted even though no significant omnibus result was found in the one-way ANOVA. This being the case, the test still would not identify any differences among means. So, in this example no differences among

means are identified in the data set in Table 1 after performing unplanned post hoc comparison tests following a non-significant omnibus F test.

INSERT TABLE 3 ABOUT HERE

However, if a priori planned comparison tests were performed in the place of the overall ANOVA, a statistically significant difference between means would be discovered. Table 3 indicates that a statistically significant ($p<0.01$) result is isolated for the hypothesis represented in coding column C5. The null hypothesis that the mean attitude-toward-school score (DV) of the two school board members equals the mean for the remaining 10 subjects is rejected (Thompson, 1987) in favor of the alternative hypothesis of difference between the two means.

Summary

Planned or a priori tests have important advantages over unplanned or post hoc tests. In addition to increasing power against Type II error rates, planned comparisons force the researcher to thoughtfully plan the research design. The example data illustrate that planned tests have more statistical power against type II error. Researchers should more often apply planned tests in their efforts to investigate important education phenomena. Appendix B summarizes the various multiple comparisons tests.

References

Fisher, R. A. (1925). Statistical methods for research workers. Edinburgh, England: Oliver and Boyd

Fisher, R. A., & Yates, F. (1957). Statistical tables for biological, agricultural and medical research (5th ed.). New York: Hafner Publishing Co.

Games, P. A. (1971a). Errata for "Multiple comparisons on means," AERJ, 1971, 531-565. American Educational Research Journal, 8, 677-678.

Games, P. A. (1971b). Multiple comparisons of means. American Educational Research Journal, 8, 531-565.

Glasnapp, D. R., & Poggio, J. P. (1985). Essentials of statistical analysis for the behavioral sciences. Columbus, OH: Merrill.

Hays, W. L. (1981). Statistics (3rd ed.). New York: Holt, Rinehart and Winston.

Horvath, T. (1985). Basic statistics for behavioral sciences. Boston: Little, Brown and Company.

Keppel, B. (1973). Design and analysis: A researcher's handbook. Englewood Cliffs, NJ: Prentice-Hall.

Keppel, G. (1982). Design and analysis: A researcher's handbook. Englewood Cliffs, NJ: Prentice-Hall.

Kirk, R. E. (1984). Elementary statistics. Monterey, CA: Brooks/ Cole.

Moore, G. W. (1983). Developing and evaluating educational research. Boston: Little, Brown and Company

Pedhazur, E. J. (1982). Multiple regression in behavioral research: Explanation and prediction (2nd ed.). New York: Holt, Rinehart and Winston.

Thompson, B. (1985). Alternate methods for analyzing data from education experiments. Journal of Experimental Education, 54, 50-55.

Thompson, B. (1987, November). The importance of a priori contrasts in analysis of variance research. Paper presented at the annual meeting of the Mid-South Educational Research Association, Mobile, AL.

Thompson, B. (1988). The importance of planned or focused comparisons in OVA research. Measurement and Evaluation in Counseling and Development, 21, 99-101.

Winer, B. J. (1971). Statistical principles in experimental design (2nd ed.). New York: McGraw-Hill.

Table 1

Hypothetical Data for Attitudes Toward School Study (n = 12)

ID	GROUP	LEVEL	DV	C1	C2	C3	C4	C5
1	Students	1	10	-1	-1	-1	-1	-1
2			20	-1	-1	-1	-1	-1
3	Teacher Aides	2	10	1	-1	-1	-1	-1
4			20	1	-1	-1	-1	-1
5	Teachers	3	10	0	2	-1	-1	-1
6			20	0	2	-1	-1	-1
7	Principals	4	10	0	0	3	-1	-1
8			20	0	0	3	-1	-1
9	Superintendents	5	10	0	0	0	4	-1
10			20	0	0	0	4	-1
11	Board Members	6	25	0	0	0	0	5
12			35	0	0	0	0	5

Table 2

One-Way ANOVA Results

SOURCE	SOS	df	MEAN SQUARE	F	p	ETA SQUARE
Between	375.000	5	75.000	1.5000	.3155	.55556
Error	300.000	6	50.000			
Total	675.000	11				

NOTE. If, in violation of conventional practice, post hoc tests (e. g., Duncan, Student-Newman-Keuls, Tukey-B, Tukey-HSD, Scheffe) were conducted in the absence of a significant omnibus result, the tests would still not identify any two groups with means significantly different at alpha equal 0.05.

Table 3

Planned Comparison Results

CONTRAST SOURCE	SOS	df	MEAN SQUARE	F	p	ETA SQUARE
C1	.000	1	.000	0.000		.0000
C2	.000	1	.000	0.000		.0000
C3	.000	1	.000	0.000		.0000
C4	.000	1	.000	0.000		.0000
C5	375.000	1	375.000	12.500	.0054	.5556
Error	300.000	6	50.000			
Total	675.000	11				

APPENDIX A

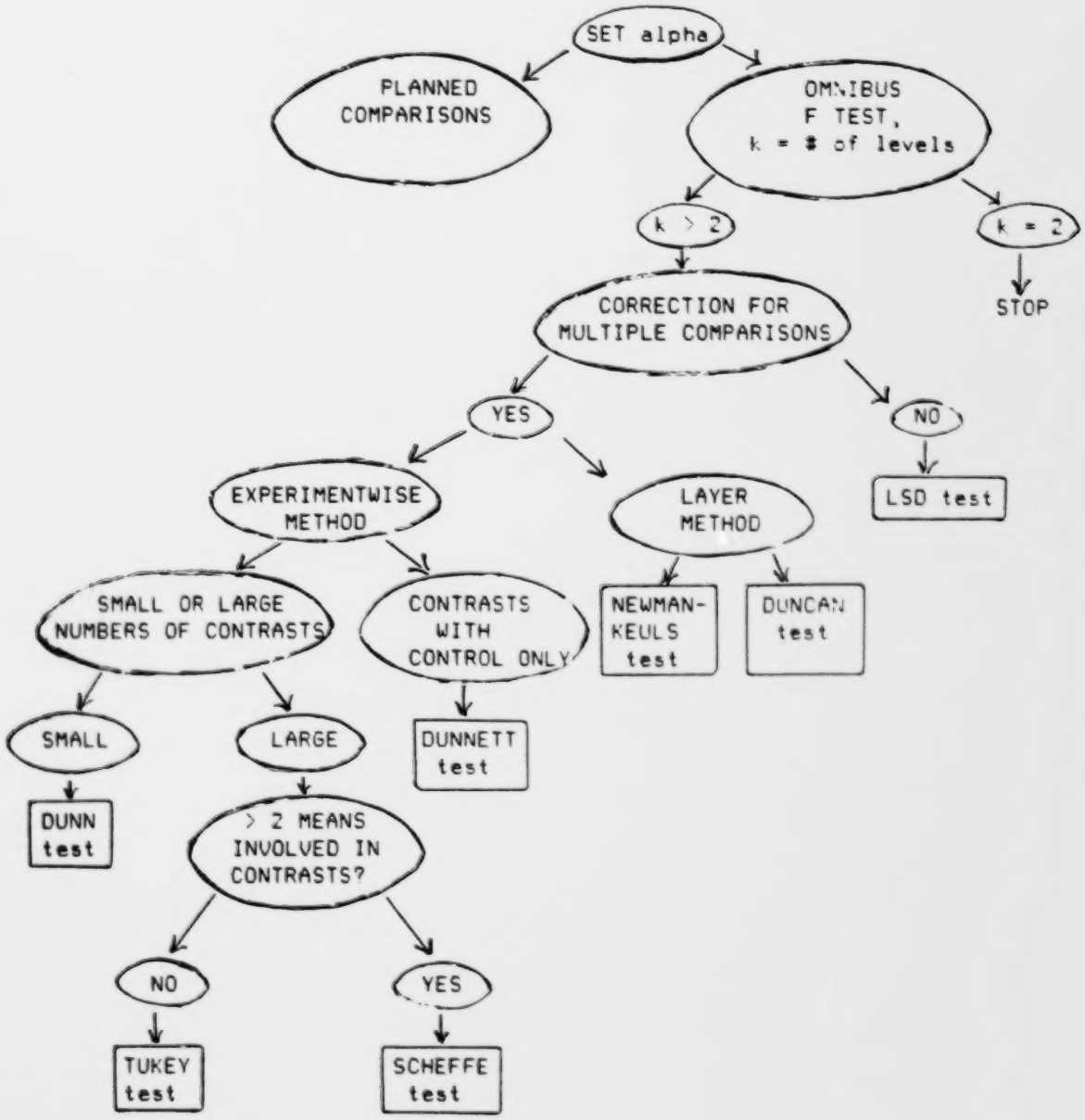
Selected SPSS-X Control Cards

```
TITLE '**********OMNIBUS no POSTHOC no  A PRIORI yes'
FILE HANDLE BT/NAME='APRIORI.DTA'
DATA LIST FILE=BT/ID 1-2 GRP 4 DV 6-7
COMPUTE C1=0
COMPUTE C2=0
COMPUTE C3=0
COMPUTE C4=0
IF (GRP LT 2)C1=-1
IF (GRP EQ 2)C1=1
IF (C1 NE 0)C2=-1
IF (GRP EQ 3)C2=2
IF (C2 NE 0)C3=-1
IF (GRP EQ 4)C3=3
IF (C3 NE 0)C4=-1
IF (GRP EQ 5)C4=4
COMPUTE C5=-1
IF (GRP EQ 6)C5=5
PRINT FORMATS C1 TO C5(F2.0)
LIST VARIABLES=ALL/CASES=12
SUBTITLE 'CONVENTIONAL ONEWAY ANOVA #1'
ONEWAY DV BY GRP(1,6)/RANGES=LSD/RANGES=DUNCAN(.05)/RANGES=SNK/
  RANGES=TUKEYB/RANGES=TUKEY/RANGES=LSDMOD/RANGES=SCHEFFE/
STATISTICS ALL
SUBTITLE 'PERFORM ANOVA USING REGRESSION #2'
REGRESSION VARIABLES=DV C1 TO C5/DESCRIPTIVES=ALL/
  DEPENDENT=DV/METHOD=STEPWISE
```

Note. Tables 1, 2, 3 and Appendix A from "The Importance of a priori Contrasts in Analysis of Variance Research," Thompson (1987).

APPENDIX B

Summary of Multiple Comparisons Tests



Note. Adapted from Keppel (1973).